

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 16, 2003, 15:50:54 : Search time 2481.53 Seconds  
(without alignments)  
16465.781 Million cell updates/sec

Title: US-09-497-967-5  
Perfect score: 1404  
Sequence: 1 atgaagaacaacatctcgtg.....tgatctcttactacctgctg 1404  
Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_ba: \*  
2: gb\_htg: \*  
3: gb\_in: \*  
4: gb\_om: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_sts: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vi: \*  
15: em\_ba: \*  
16: em\_fun: \*  
17: em\_hum: \*  
18: em\_in: \*  
19: em\_mu: \*  
20: em\_om: \*  
21: em\_or: \*  
22: em\_ov: \*  
23: em\_pat: \*  
24: em\_ph: \*  
25: em\_pl: \*  
26: em\_ro: \*  
27: em\_sts: \*  
28: em\_un: \*  
29: em\_vi: \*  
30: em\_htg\_hum: \*  
31: em\_htg\_inv: \*  
32: em\_htg\_other: \*  
33: em\_htg\_mus: \*  
34: em\_htg\_pln: \*  
35: em\_htg\_rod: \*  
36: em\_htg\_mam: \*  
37: em\_htg\_vrt: \*  
38: em\_sy: \*  
39: em\_htgo\_hum: \*  
40: em\_htgo\_mus: \*  
41: em\_htgo\_other: \*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	782.6	55.7	3026	3	AF324424	AF324424 Ichthyoph
2	63.2	4.5	2486	3	AF140273	AF140273 Ichthyoph
3	55.2	3.9	125020	9	AF429315	AF429315 Homo sapi
4	54.4	3.9	1520	3	AF405431	AF405431 Ichthyoph
5	52.2	3.7	1249	3	ICYIMANT	M92907 Ichthyophth
6	50	3.6	125020	9	AF429315	AF429315 Homo sapi
7	43.4	3.1	230372	2	AC073693	AC073693 Mus muscu
8	41.4	2.9	159764	9	AC093866	AC093866 Homo sapi
9	41	2.9	643	8	AF374493	AF374493 Thalassio
10	41	2.9	179217	2	AC099748	AC099748 Bos tauru
11	39.4	2.8	591	8	AF374552	AF374552 Thalassio
12	38.6	2.7	643	8	AF374492	AF374492 Bos tauru
13	38.6	2.7	16814	9	HSCOLPDGF	AF374492 Thalassio
14	38.6	2.7	341950	1	AP003600	X8705 H.sapiens D
15	38.4	2.7	224271	9	AC026803	X8705 H.sapiens D
16	38.2	2.7	15294	1	AE006936	AE006936 Mycobacte
17	38.2	2.7	31300	1	MTV035	AC026803 Homo sapi
18	38.2	2.7	152777	2	AC125855	AC026803 Homo sapi
19	38.2	2.7	233345	4	AJ421481	AC125855 Rattus no
20	38.2	2.7	338150	1	AP005278	AC125855 Rattus no
21	38.2	2.7	349980	6	AX127147	AJ421481 Bos tauru
22	37.8	2.7	559	8	AF374543	AP005278 Corynebac
23	37.8	2.7	591	8	AF374540	AX127147 Sequence
24	37.8	2.7	591	8	AF374542	AF374543 Thalassio
25	37.8	2.7	591	8	AF374544	AF374540 Thalassio
26	37.8	2.7	591	8	AF374545	AF374542 Thalassio
27	37.8	2.7	591	8	AF374547	AF374544 Thalassio
28	37.8	2.7	591	8	AF374548	AF374545 Thalassio
29	37.8	2.7	591	8	AF374549	AF374547 Thalassio
30	37.8	2.7	591	8	AF374550	AF374548 Thalassio
31	37.8	2.7	591	8	AF374551	AF374549 Thalassio
32	37.8	2.7	643	8	AF374490	AF374551 Thalassio
33	37.8	2.7	643	8	AF374491	AF374550 Thalassio
34	37.8	2.7	643	8	AF374492	AF374490 Thalassio
35	37.8	2.7	643	8	AF374493	AF374491 Thalassio
36	37.8	2.7	643	8	AF374494	AF374492 Thalassio
37	37.8	2.7	643	8	AF374495	AF374493 Thalassio
38	37.8	2.7	643	8	AF374496	AF374494 Thalassio
39	37.8	2.7	643	8	AF374497	AF374495 Thalassio
40	37.8	2.7	643	8	AF374498	AF374496 Thalassio
41	37.8	2.7	643	8	AF374499	AF374497 Thalassio
42	37.8	2.7	643	8	AF374500	AF374498 Thalassio
43	37.8	2.7	643	8	AF374501	AF374499 Thalassio
44	37.8	2.7	643	8	AF374502	AF374500 Thalassio
45	37.8	2.7	643	8	AF374503	AF374501 Thalassio
46	37.8	2.7	643	8	AF374504	AF374502 Thalassio
47	37.8	2.7	643	8	AF374505	AF374503 Thalassio
48	37.8	2.7	643	8	AF374506	AF374504 Thalassio
49	37.8	2.7	643	8	AF374507	AF374505 Thalassio
50	37.8	2.7	643	8	AF374508	AF374506 Thalassio
51	37.8	2.7	643	8	AF374509	AF374507 Thalassio
52	37.8	2.7	643	8	AF374510	AF374508 Thalassio
53	37.8	2.7	643	8	AF374511	AF374509 Thalassio
54	37.8	2.7	643	8	AF374512	AF374510 Thalassio
55	37.8	2.7	643	8	AF374513	AF374511 Thalassio
56	37.8	2.7	643	8	AF374514	AF374512 Thalassio
57	37.8	2.7	643	8	AF374515	AF374513 Thalassio
58	37.8	2.7	643	8	AF374516	AF374514 Thalassio
59	37.8	2.7	643	8	AF374517	AF374515 Thalassio
60	37.8	2.7	643	8	AF374518	AF374516 Thalassio
61	37.8	2.7	643	8	AF374519	AF374517 Thalassio
62	37.8	2.7	643	8	AF374520	AF374518 Thalassio
63	37.8	2.7	643	8	AF374521	AF374519 Thalassio
64	37.8	2.7	643	8	AF374522	AF374520 Thalassio
65	37.8	2.7	643	8	AF374523	AF374521 Thalassio
66	37.8	2.7	643	8	AF374524	AF374522 Thalassio
67	37.8	2.7	643	8	AF374525	AF374523 Thalassio
68	37.8	2.7	643	8	AF374526	AF374524 Thalassio
69	37.8	2.7	643	8	AF374527	AF374525 Thalassio
70	37.8	2.7	643	8	AF374528	AF374526 Thalassio
71	37.8	2.7	643	8	AF374529	AF374527 Thalassio
72	37.8	2.7	643	8	AF374530	AF374528 Thalassio
73	37.8	2.7	643	8	AF374531	AF374529 Thalassio
74	37.8	2.7	643	8	AF374532	AF374530 Thalassio
75	37.8	2.7	643	8	AF374533	AF374531 Thalassio
76	37.8	2.7	643	8	AF374534	AF374532 Thalassio
77	37.8	2.7	643	8	AF374535	AF374533 Thalassio
78	37.8	2.7	643	8	AF374536	AF374534 Thalassio
79	37.8	2.7	643	8	AF374537	AF374535 Thalassio
80	37.8	2.7	643	8	AF374538	AF374536 Thalassio
81	37.8	2.7	643	8	AF374539	AF374537 Thalassio
82	37.8	2.7	643	8	AF374540	AF374538 Thalassio
83	37.8	2.7	643	8	AF374541	AF374539 Thalassio
84	37.8	2.7	643	8	AF374542	AF374540 Thalassio
85	37.8	2.7	643	8	AF374543	AF374541 Thalassio
86	37.8	2.7	643	8	AF374544	AF374542 Thalassio
87	37.8	2.7	643	8	AF374545	AF374543 Thalassio
88	37.8	2.7	643	8	AF374546	AF374544 Thalassio
89	37.8	2.7	643	8	AF374547	AF374545 Thalassio
90	37.8	2.7	643	8	AF374548	AF374546 Thalassio
91	37.8	2.7	643	8	AF374549	AF374547 Thalassio
92	37.8	2.7	643	8	AF374550	AF374548 Thalassio
93	37.8	2.7	643	8	AF374551	AF374549 Thalassio
94	37.8	2.7	643	8	AF374552	AF374550 Thalassio
95	37.8	2.7	643	8	AF374553	AF374551 Thalassio
96	37.8	2.7	643	8	AF374554	AF374552 Thalassio
97	37.8	2.7	643	8	AF374555	AF374553 Thalassio
98	37.8	2.7	643	8	AF374556	AF374554 Thalassio
99	37.8	2.7	643	8	AF374557	AF374555 Thalassio
100	37.8	2.7	643	8	AF374558	AF374556 Thalassio

ALIGNMENTS

RESULT 1  
AF324424  
LOCUS  
DEFINITION  
AF324424  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS

AF324424  
Ichthyophthirius multifiliis  
(IAG52A) gene, complete cds.  
AF324424  
GI:12698726  
Ichthyophthirius multifiliis.  
Ichthyophthirius multifiliis  
Eukaryota; Alveolata; Ciliophora; Oligohymenophorea;  
Hymenostomatida; Ophryoglenina; Ichthyophthirius.  
1 (bases 1 to 3026)  
Lin, Y., Lin, T. L., Wang, C. C., Wang, X., Stieger, K., Klopfleisch, R.  
and Ciark, T. G.

3026 bp DNA linear INV 27-FEB-2002  
Ichthyophthirius multifiliis immobilization antigen isoform

[illegible]

DEFINITION Ichthyophthirius multifiliis immobilization antigen precursor (IAG48) gene, complete cds.  
ACCESSION AF140273  
VERSION AF140273.1 GI:4868370  
KEYWORDS  
SOURCE  
ORGANISM Ichthyophthirius multifiliis.  
Ichthyophthirius multifiliis.  
Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomata; Ophryoglenina; Ichthyophthirius.  
REFERENCE 1 (bases 1 to 2486)  
AUTHORS Clark, T.G., Lin, T.-L., Jackwood, D.A., Sherrill, J., Lin, Y. and Dickerson, H.W.  
TITLE The gene for an abundant parasite coat protein predicts tandemly repetitive metal binding domains  
JOURNAL Gene 229 (1-2), 91-100 (1999)  
MEDLINE 99196987  
PUBMED 10095108  
REFERENCE 2 (bases 1 to 2486)  
AUTHORS Gaertig, J., Gao, Y., Tishgarten, T., Clark, T.G. and Dickerson, H.W.  
TITLE Surface display of a parasite antigen in the ciliate Tetrahymena thermophila  
JOURNAL Nat. Biotechnol. (1999) In press  
REFERENCE 3 (bases 1 to 2486)  
AUTHORS Clark, T.G., Lin, T.-L., Jackwood, D.A. and Dickerson, H.W.  
TITLE Direct Submission  
JOURNAL Submitted (29-MAR-1999) Microbiology & Immunology, Cornell University, College of Veterinary Medicine, Ithaca, NY 14853, USA  
FEATURES  
source  
/organism="Ichthyophthirius multifiliis"  
/serotype="A"  
/isolate="GI"  
/db\_xref="taxon:5932"  
399..1845  
/gene="IAG48"  
399..1845  
/gene="IAG48"  
/product="immobilization antigen precursor"  
433..1761  
/note="surface"  
/name="IAG48"  
/codon\_start=1  
/transl\_table=6  
/product="immobilization antigen precursor"  
/protein\_id="AAD31283.1"  
/db\_xref="GI:4868371"  
/translation="MKYNILLIILISLFINELRAVPCPDGTQTQAGLDVGAADIGTC VNCRPNFYNGAAGEANGNPFAANNAARQICVPCQINRVGSVTNAGDLATATOC STQCTGTALDDGTDVDFRSAAQCVCVKPNFYNGGSPQGPARGVQVFAAGAAAGV AAVTSCVPCQLKNDSPATAGANLATQCSNQCPTGTVLDDGTVLFTSATLGVK CRNFYNGGSPQGPARGVQVFAAGAAAGVAVTSCVPCQINKNDSPTAGAAQNL ATCSTQCPTGTATQDQVLYVSNSTQCSQCIANYFFNGFPAQKSQLKPCVSKTT PAHAPNTATQATQCLTTCPAGTVLDDGTSITNFVASATECKCSAGFFASKTTGTAG TDTCTCTKRLTSGATAKYVAEATQKQCASTTFKFLSILLSFTSYLL"  
433..492  
sig\_peptide  
/gene="IAG48"  
493..1758  
mat\_peptide  
/gene="IAG48"  
/product="immobilization antigen"

BASE COUNT 896 a 310 c 321 g 959 t  
ORIGIN  
Query Match 4.5%; Score 63.2; DB 3; Length 2486;  
Best Local Similarity 52.2%; Pred. No. 5 6e-07;  
Matches 140; Conservative 0; Mismatches 12; Indels 0; Gaps 0;  
QY 1098 TACCGCTGACCGAGACCGCTACCTGATCGCTCAGTGTGCTCGAGTGTCTCTGCTGCGAAC 1157  
Db 1443 TGCCTCAGGTAATACGTCTACTTTAACCCACATAATGTTTGACACATGCTCTGCTGGTAC 1502  
QY 1158 CGTCTGACCGACCGACCGACCTCTACCTACAAGCAGCGCTGCTTCTGAGTGTGGAAGTG 1217  
Db 1503 AGTACTTGATGATGACATCAACTAATTTTCTAGCTCCCGCACTGAATGTACTAATATG 1562

QY 1218 TGCTGCTAACTTCTACACCACGACGACCGACTGGTGGTGAATCGACACCTGTATC 1277  
Db 1563 TTCTGCTGGCTTTTTCATCAAAAACAACTGGTTTACAGCAGGTACTGTACATGTAC 1622  
QY 1278 CTCTTGTAAAGAAGCTGACCTCTGGAGCTGAGCTTAACCTCGCTGAGCTGCTAAGAA 1337  
Db 1623 TGAATGTAC/AAAAAATAACTTCTGTCGCCACAGCTAAAGTATATGCTGAAGCTACTCA 1682  
QY 1338 GAACATCCAGTGTGACTTTCCTCAACTTC 1365  
Db 1683 AAAATATATATGCGCTCCCACTACTTTC 1710  
RESULT 3  
AF429315  
LOCUS AF429315 125020 bp DNA linear PRI 18-JAN-2002  
DEFINITION Homo sapiens junctophilin 3 (JPH3) gene, partial cds.  
ACCESSION AF429315  
VERSION AF429315.1 GI:17646244  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 125020)  
AUTHORS Holmes, S.E., O'Hearn, E., Rosenblatt, A., Callahan, C., Hwang, H.S., Ingersoll-Ashworth, R.G., Fleisher, A., Stevanin, G., Brice, A., Potter, N.T., Ross, C.A. and Margolis, R.L.  
TITLE A repeat expansion in the gene encoding junctophilin-3 is associated with Huntington disease-like 2  
JOURNAL Nat. Genet. 29 (4), 377-378 (2001)  
MEDLINE 21583737  
PUBMED 11694876  
REFERENCE 2 (bases 1 to 125020)  
AUTHORS Holmes, S.E., Ingersoll-Ashworth, R.G., Ross, C.A. and Margolis, R.L.  
TITLE Direct Submission  
JOURNAL Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA  
FEATURES  
source  
1..125020  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/map="16q24.3; between D16S520 and WI-12410"  
/note="isolated from a patient with Huntington's Disease-Like 2 (HDL2)"  
complement(35581..35746)  
/rpt\_type=tandem  
/rpt\_unit=ctg  
complement(<36507..>36887)  
/gene="JPH3"  
/note="JPH3"  
complement(<36507..>36887)  
/gene="JPH3"  
/product="junctophilin 3"  
complement(<36507..>36887)  
/gene="JPH3"  
/note="component of the junctional complex between plasma membrane and endoplasmic reticulum"  
/codon\_start=1  
/product="junctophilin 3"  
/protein\_id="AAL40941.1"  
/db\_xref="GI:17646245"  
/translation="MSSGGRFNFDGSGYCGWEDGKAHGVCTGPKGGEYTGNS HGFVGLVYTFPSGNTYQGTWAQRHIGLESKRWYKGTGTHGFKRGYVRECA G NAKYEGTWSNGLQDQYGTETYSQ"

BASE COUNT 29056 a 32731 c 30696 g 28283 t 4254 others  
ORIGIN  
Query Match 3.9%; Score 55.2; DB 9; Length 125020;  
Best Local Similarity 10.9%; Pred. No. 0.00017;  
Matches 108; Conservative 421; Mismatches 452; Indels 6; Gaps 3;



```

RESULT 5
ICYMANT
LOCUS      1249 bp      mRNA      linear      INV 18-SEP-1998
DEFINITION Ichthyophthirius multifiliis immobilization antigen precursor,
            mRNA, partial cds.
ACCESSION  M92907
VERSION     M92907.1 GI:3628568
SOURCE      Ichthyophthirius multifiliis.
            Ichthyophthirius multifiliis.
            Eukaryota; Alveolata; Ciliophora; Oligohymenophorea;
            Hymenostomatida; Ophryoglenina; Ichthyophthirius.
            1 (bases 1 to 1249)
REFERENCE   Lin,T.L. and Dickerson,H.W.
            Purification and partial characterization of immobilization
            antigens from Ichthyophthirius multifiliis
            J. Protozool. 39 (4), 457-463 (1992)
JOURNAL     93020550
MEDLINE     1383510
PUBMED      1383510
REFERENCE   2 (bases 1 to 1249)
AUTHORS     Clark,T.G., McGraw,R.A. and Dickerson,H.W.
TITLE       Developmental expression of surface antigen genes in the parasitic
            ciliate Ichthyophthirius multifiliis
JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 89 (14), 6363-6367 (1992)
MEDLINE     92335298
PUBMED      1631132
REFERENCE   3 (bases 1 to 1249)
AUTHORS     Clark,T.
TITLE       Direct Submission
JOURNAL     Submitted (06-JUN-1992) Microbiology and Immunology, Cornell
            University, Ithaca, NY 14853, USA
REFERENCE   4 (bases 1 to 1249)
AUTHORS     Clark,T.
TITLE       Direct Submission
JOURNAL     Submitted (18-SEP-1998) Microbiology and Immunology, Cornell
            University, Ithaca, NY 14853, USA
REMARK      Sequence update by submitter
COMMENT      On Sep 18, 1998 this sequence version replaced gi:159289.
FEATURES    Location/Qualifiers
             source
               1..1249
               /organism="Ichthyophthirius multifiliis"
               /strain="Georgia"
               /db_xref="taxon:5932"
               /dev_stage="trophont"
               <1..1190
               /note="48 kDa"
               /citation=[2]
               /citation=[1]
               /codon_start=3
               /evidence=experimental
               /product="immobilization antigen precursor"
               /protein_id="AAC36158.1"
               /db_xref="GI:3628569"
               /translation="AVPCPDGTQTQAGLTDVGAADLGTVCNCRPNFYNGSAAQGEAN
               GNQFAANNAARGICVPCQINRVGSVTNAGDLATLQCSQCTGTALDDGVTDVFD
               RSAAQCVKCKPNFYNGSGPOGEARGVQVFAAGAAAGVAAVTSQCVPQCNKNDSPA
               TAGAANLATQCSNCGCTPTGLDGVTLVFNFSATLVCVKRPNFYNGSGPOGEARGV
               QVFAGAAAGAAVATISQCVPCQINKNDSPTAGAAANLATQCSQCTGTATQDQVTV
               LVFNSSTQCSQCIANYEFNGLKAGSKQLKCPVSKTTPAHAPGNATQATQCLTTC
               PAGTVLDGDTSTNFVASATECTKCSAGFFASKTTGTAGTDTCTECTKLTSGATAVN
               IHQ"
             slg_peptide <1..5
             mat_peptide 6..1187
             repeat_region /product="immobilization antigen"
             /evidence=experimental
             177..893
             /citation=[2]
             /rpt_type=tandem
             /rpt_unit=414..680
             BASE COUNT 348 a 243 c 257 g 401 t
ORIGIN

```

```

Query Match      3.7%; Score 52.2; DB 3; Length 1249;
Best Local Similarity 52.5%; Pred. No. 0.00089;
Matches 114; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 1098 TACCGCTGGAGGAACCGCTACCCCTGATCGCTCAGTGTGCTCTGGAGTGTCTGCTGCTGGAAC 1157
DB      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
956 TGCTCCAGGTAATCTGCTTACTTAAGCCACATAATGTTGACCACATGCTCTGCTGGTAC 1015

QY 1158 CGTGTGACCGGACGACACCTCTACCTACAGCAGGCTGCTTCTGTAGTGTGTAAGTG 1217
DB      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1016 AGTACTTGTATGATGAACATCACTAAATTTGTAGCTTCCGCACTGAATGTACTAAATG 1075

QY 1218 TGTCTGCTAACTTCTACACCACCAAGCAGACCTGGGTGCTGGAATCGACACCTGTAC 1277
DB      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1076 TTCTGCTGCTTTTGTGATCAAAAACACCTGGTTTACACGACGCTACTGATACATGTAC 1135

QY 1278 CTCCTGTACACAGAGCTGACCTCTGGAGCTGAGGCT 1314
DB      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1136 TGAATGTACTAAAAAATTAACCTTCTGTCACACAGCT 1172

RESULT 6
AF429315/C
LOCUS      Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
DEFINITION AF429315
ACCESSION  AF429315
VERSION     AF429315.1 GI:17646244
KEYWORDS
SOURCE      Homo sapiens.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 125020)
REFERENCE   Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,
            Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,
            Potter,N.T., Ross,C.A. and Margolis,R.L.
            A repeat expansion in the gene encoding junctophilin-3 is
            associated with Huntington disease-like 2
            Nat. Genet. 29 (4), 377-378 (2001)
JOURNAL     21583737
MEDLINE     11694876
PUBMED
REFERENCE   2 (bases 1 to 125020)
AUTHORS     Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.
TITLE       Direct Submission
JOURNAL     Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical
            Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
FEATURES    Location/Qualifiers
             source
               1..125020
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /chromosome="16"
               /map="16q24.3; between D16S520 and WI-12410"
               /note="Isolated from a patient with Huntington's
               Disease-Like 2 (HDL2)."
               complement(35581..35746)
               /rpt_type=tandem
               /rpt_unit=ctg
               complement(<36507..>36887)
               /gene="JPH3"
               /note="Jp3"
               complement(<36507..>36887)
               /gene="JPH3"
               /product="Junctophilin 3"
               complement(<36507..36887)
               /gene="JPH3"
               /note="component of the junctional complex between plasma
               membrane and endoplasmic reticulum"
               /codon_start=1
               /product="Junctophilin 3"
               /protein_id="AAL40941.1"
               /db_xref="GI:17646245"
               /translation="MSSGGRFNFDGSGYCGGWEDGKAHGHCYCTGPKGGEVYTGSMW
               HGFEVLGVYTPSGNTYQGTWAOCKRHGIGLESKGWYKGVTHGFKRGYRGVACAG

```

Thu Feb 20 11:10:35 2003

BASE COUNT	29056 a	32731 c	30696 g	28283 t	4254 others
ORIGIN	NGAKEYGTWNGLDQGYGTETYSDE**				
Query Match	3.6%; Score 50; DB 9; Length 125020;				
Best Local Similarity	10.1%; Pred. No. 0.0056;				
Matches	93;	Conservative	409;	Mismatches	406; Indels 17; Gaps 3;
Qy	274	AAGCTGAAGTGTCTCGCTGGAAACCGCTATCGCTGGAGGAGCTACCGACTACGGCTGCATC	333		
Db	17784	AGKSRGRTGYMKKGGKSMSSKWKGSSTSRRRSSAKSCSYMWGMSMCMSCMSM	17725		
Qy	334	ATCACCGAGTGTGAAGTGTGCGATCACTTCTACACGAGAGAGCTCTCAATCTTAAC	393		
Db	17724	AKSYMYCYMYRMSYMYYYKCYSCMGMSSTSYSCWKMSWGSYCKMYIYSGWS	17665		
Qy	394	GCTGGAGCTTCTACCTGTACCGCTGTCTGTGAACCGCTGGGAGGAGCTCT-----GA	448		
Db	17664	SYSTSMGYSSSTCKKIKCSWSMYKCKTSKYRKRYSYYWGGKRAKKYKYGAGRRR	17605		
Qy	449	CCGCTGGAACGCTGCTACCATCGTGGCTCAGTGTAACTGTAAGTGTGCTTCCACCGAACCG	508		
Db	17604	MSYWKCAKWWMSYCCWSYCMYTSKCTYKSKYRGYYGSKTCYSAGGKSRMY	17545		
Qy	509	CTCTGGAGCAGGAGTGACCGGACTAGCTGGCTCTTCAACCGAGT-GTGTGAAGTGT	567		
Db	17544	YCMRRSSKSSWSMARSWCMGWAGYRRSKRAGWRRSKGRSTGMKRAKCSK	17485		
Qy	568	CGCTGAAGTCTTCTACTACACGGAACAAACGGAACACCCCTTCAACCCCTGGAAAGTCT	627		
Db	17484	TGSYGSTGRSMKKKGYSKYSRGMKKKTKCYCMWYKYRKTSCWYMYKMSWGYKRYK	17425		
Qy	628	CAGTGTACCCCTGTCTCTGTATCAAGCTGTAGCTGGCTCAGGCTACCCCTGGGAAC	687		
Db	17424	RCCMKKKGCTGYRGMSSKSGYSMGRSSYSTSCWWSWGYSMWKMKYNSYKRRR	17365		
Qy	688	GACCTACCATCAGCTCAGTGTAGCTGGCTGTGCTTCAAGGGAACCATCTCTCTGCT	747		
Db	17364	SMRGSMSKRWGYAGRCYSSSWSTRKRSKCYKSYKKGKRGKMGKMGKRGSKYW	17305		
Qy	748	GGAGTGAACACTGGGTGGGTGACACACGAGTGTACCAACTGTCTCTCAACTCTAC	807		
Db	17304	SMKMKRRSSMKYSTKYSKGRRRSKGWGRSTKSKAKSSMRMAGSKCTYGSYSWNNR	17245		
Qy	808	AACAACACGCTCTAATCAACCCCTGGAAACTCTACCTCTGCTGCTCTCTCTGCTAAC	867		
Db	17244	RNNR-----MKTGCNMYRBRAMNGNAAAGCTTCCCCANTNGGGGAAAA	17196		
Qy	868	AAGGACTACGAGGTGAGGCTACCGCTGGAGGAGCTGCTACCTGGCTAAGCAGTGAAC	927		
Db	17196	AGGGGGSASRASCYKGRMSKSYRSRGTTRRCMKSKORRSGKSGMTGRSGGKTSYS	17136		
Qy	928	ATCGCTTCTCTGAGGAAACCGCTATCGCTTCTGAGCTACCAACTACCTGATCTCTGCG	987		
Db	17136	AKSGRGCTGYCWGKRGKCKMSRKKMKYKSYRERHWTCKMKCYMYMYMYCRSMCC	17076		
Qy	988	ACCGAGTGTGAAGTGTGCTGCTAATCTTCTTCTGAGGAAACAACTTCCAGGTGGA	1047		
Db	17076	MCWKSCCCGYCMGMSYSYSGYKSWGKSYMYRYSYKRSKRSKAWRSRGRMGTOGRYK	17016		
Qy	1048	TCTTCTCGCTGAAGCTGTCTCTGCTTCTAACAAGGTGCGAGGAGCTGTGGCTACCGCTG	1107		
Db	17016	GGRSYKGGGYSYKGGGSGWGGKGRSGSAGKRSYKMKSSCARYSKYSKYSRW	16956		
Qy	1108	GGAACCGCTACCGCTGATCGCTCAGTGTGCTTCTGAGGTGTCTGCTGGAACCGCTGAC	1167		
Db	16956	WMSYYYCWGKCSWSSWSSWSSWSSWSSWSSWSSWSSWSSWSSWSSWSSWSSWSS	16896		
Qy	1168	GAGGAAACCATCTCTACCTACAGC	1192		
Db	16896	KWSMRGRYYCTDSMKMYITGC	16871		

RESULT 7  
AC073693/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

AC073693 230372 bp DNA linear HTG 29-JUN-2000  
Mus musculus clone RP23-152L22, WORKING DRAFT SEQUENCE, 19  
unordered pieces.  
AC073693  
AC073693.1 GI:8810310  
HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
Mus musculus.  
Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 230372)  
DOE Joint Genome Institute.  
Sequencing of Mouse  
Unpublished  
2 (bases 1 to 230372)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov  
-----  
Project Information  
Center Project Name: 1795526  
Center clone name: RPCI-23\_152L22  
-----  
Summary Statistics  
Consensus quality: 217356 bases at least Q40  
Consensus quality: 223517 bases at least Q30  
Consensus quality: 224719 bases at least Q20  
Estimated insert size: 258000; adarose-fp estimation  
Estimated insert size: 228572; sum-of-contigs estimation  
Quality coverage: 7.65 in Q20 bases; agarose-fp estimation  
Quality coverage: 8.64 in Q20 bases; sum-of-contigs estimation.  
NOTE: This is a 'working draft' sequence. It currently  
\* consists of 19 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1087: contig of 1087 bp in length  
\* 1088  
\* 1187: gap of unknown length  
\* 1188  
\* 2715: contig of 1528 bp in length  
\* 2716  
\* 2815: gap of unknown length  
\* 2816  
\* 4158: contig of 1343 bp in length  
\* 4159  
\* 4258: gap of unknown length  
\* 4259  
\* 6473: contig of 2215 bp in length  
\* 6474  
\* 6573: gap of unknown length  
\* 6574  
\* 8982: contig of 2409 bp in length  
\* 8983  
\* 9082: gap of unknown length  
\* 11186: contig of 2104 bp in length  
\* 11187  
\* 11287: gap of unknown length  
\* 17000: contig of 5714 bp in length  
\* 17001  
\* 17101: gap of unknown length  
\* 17101  
\* 25179: contig of 8079 bp in length  
\* 25180  
\* 25279: gap of unknown length  
\* 34875: contig of 9596 bp in length  
\* 34876  
\* 34876: gap of unknown length  
\* 34876  
\* 43664: contig of 8689 bp in length  
\* 43665  
\* 43665: gap of unknown length  
\* 53788: contig of 10024 bp in length  
\* 53789  
\* 53888: gap of unknown length  
\* 53889  
\* 65520: contig of 11632 bp in length  
\* 65521  
\* 80284: gap of unknown length  
\* 80284: contig of 14664 bp in length  
\* 80285  
\* 80384: gap of unknown length  
\* 80385  
\* 98449: contig of 18065 bp in length

```

* 98450 98549: gap of unknown length
* 98550 116262: contig of 17713 bp in length
* 116263 116362: gap of unknown length
* 116363 140706: contig of 24344 bp in length
* 140707 140806: gap of unknown length
* 140807 167609: contig of 26803 bp in length
* 167610 167709: gap of unknown length
* 167710 199398: contig of 31689 bp in length
* 199399 199498: gap of unknown length
* 199499 230372: contig of 30874 bp in length.
FEATURES
    source
        Location/Qualifiers
            1..230372
                /organism="Mus musculus"
                /db_xref="taxon:10090"
                /clone="RP23-152L22"
                /clone_lib="RPCI mouse BAC library 23"
BASE COUNT 61266 a 53740 c 52522 g 61042 t 1802 others
ORIGIN

```

```

Query Match      3.1%; Score 43.4; DB 2; Length 230372;
Best Local Similarity 53.9%; Pred. No. 0.51;
Matches 89; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 166 GCTGCTGCTTCGCTGGAGCTCTACCTGTACCCCTTCCTCAGAGAAGACGCT 225
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 212511 GCTGCTGCTCCCTGGCTGCTGCTCCTCCCTGGCTGCTGCTGCTTCCTCCCTGGCTGCT 212452

QY 226 GGAGCTGACCGTAACCCCTGCTACCGCTAACCTGGTGACCCAGTGTACGTCGAAGTGT 285
      ||| || ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 212451 GCTGCTGCTCCCTGGCTGCTGCTACTGCTCCCTGGATGCTGCTGCTCCCTGAATGGT 212392

QY 286 CCTGCTGGAACCGCTATCGCTGGAGGAGCTACCGACTACGCTGCT 330
      ||||| || ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 212391 GCTACTGCTCCCTGGCTGCTGCTGCTGCTCCTCCCTGGATGCTGCT 212347

```

```

RESULT 8
AC093866
LOCUS      AC093866             159764 bp    DNA    linear    PRI 01-MAR-2002
DEFINITION Homo sapiens BAC clone RP11-582C12 from 4, complete sequence.
ACCESSION  AC093866 AC025608
VERSION    AC093866.3 GI:16973777
KEYWORDS   HTS.
SOURCE     Homo sapiens.
ORGANISM   Homo sapiens.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            Sulston,J.E. and Waterston,R.
            Toward a complete human genome sequence
            Genome Res. 8 (11), 1097-1108 (1998)
            99063792
            9847074
REFERENCE  2 (bases 1 to 159764)
            Pearman,C., Meyer,R. and Doeberber,A.
            The sequence of Homo sapiens BAC clone RP11-582C12
            Unpublished (2001)
REFERENCE  3 (bases 1 to 159764)
            Waterston,R.H.
            Direct Submission
            Submitted (10-SEP-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
            4 (bases 1 to 159764)
            Waterston,R.H.
            Direct Submission
            Submitted (17-NOV-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
            5 (bases 1 to 159764)
            Waterston,R.H.
            Direct Submission
            Submitted (03-JAN-2002) Genome Sequencing Center, Washington

```

```

University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
6 (bases 1 to 159764)
Waterston,R.
Direct Submission
Submitted (01-MAR-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Nov 17, 2001 this sequence version replaced gi:15987350.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
Center project name: H_NH0582C12
Drafting Center: WIBR
-----

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tatenno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pletter de Jong and coworkers at <http://www.chori.org>

VECTOR: pBACe3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-549C16, 2000 bp overlap; the clone sequenced to the right is RP11-115D19. Actual start of this clone is at base position 123633 of RP11-549C16; actual end is at base position 159764 of RP11-582C12.

Sequence derived from one plasmid subclone, base position 55515 to 55572.

The sequence of AC025608 has been incorporated into AC093866.

```

FEATURES
    source
        Location/Qualifiers
            1..159764
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="4"
                /map="4"
                /clone="RP11-582C12"
                /clone_lib="RPCI-11"
                /rpt_family="ERV1"
                /rpt_family="ERV1"
                /rpt_family="ERV1"
                /rpt_family="ERV1"
                /rpt_family="MaLR"
            repeat_region
            repeat_region
            repeat_region
            repeat_region

```





```
/product="sexually induced protein SIG 1"
/protein_id="AAK61135.1"
/db_xref="GI:14336350"
/translation="SHDMCTCDNRNFGQADCSLRTPFGKAHVDPGRGLDASLTIGDH
YDIAAGQSLPYGITEGFPMSDGTAGTAVTANTAHYMECSNKGICDRINVLCPCPG
YDGAACQASRCSKANSSTPGSGGERSNTNFKVENASAFHGRAAEVQVQDCQSGHG
TCMTIEQLAFLDHGNTYDLWDKDTM"
BASE COUNT 166 a 163 c 144 g 170 t
ORIGIN

Query Match 2.9%; Score 41; DB 8; Length 643;
Best Local Similarity 55.2%; Pred. No. 1.6;
Matches 80; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 278 TGAAGTGTCTCTGCTGGAACCGCTATCGCTGGAGGAGCTACGACTACGCTCTATCATCA 337
Db 282 TGTGGATACTGCTGGAACCGTATCGCTACACGGCTCATGACTACATGGAATGTTCCA 341
QY 338 CCGAGTGTGAACTGTGCGATCACTTCTACAACGAGACGCTCCTCAACTTCAACGCTG 397
Db 342 ATAAGGATATATGCGATCGCATCATGCTCTTGGGAATGTCCTCGGCTATCACGGTG 401
QY 398 GAGCTTCTACCTGTACCGCTGTCC 422
Db 402 CAGCTGCCAGGCTCATCATGTCC 426

RESULT 10
AC099748
LOCUS
DEFINITION Bos taurus clone RP42-504H16, WORKING DRAFT SEQUENCE, 5 ordered
pieces.
ACCESSION AC099748
VERSION AC099748.2 GI:21427692
KEYWORDS HTGS_PHASE2; HTGS_DRAFT.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 179217)
Akhter, N., Antonellis, A., Ayele, K., Beckstrom-Sternberg, S.M.,
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C.,
Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J.,
Haghighi, P., Hansen, N., Ho, S.-L., Idol, J.R., Karlins, E., Laric, P.,
Lee-Lin, S.-Q., Legaspi, R., Maduro, Q.L., Maduro, V.B.,
Margulies, E.H., Masiello, C., Maskeri, B., Mastrian, S.D.,
McCloskey, J.C., McDowell, J., Paguirigan, C., Pearson, R.,
Portnoy, M.E., Prasad, A., Schueler, M.G., Stantripop, S., Thomas, J.W.,
Thomas, P.J., Touchman, J.W., Tsurgueon, C., Vogt, J.L., Walker, M.A.,
Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 179217)
Green, E.D.
Direct Submission
Submitted (20-NOV-2001) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA
3 (bases 1 to 179217)
Green, E.D.
Direct Submission
Submitted (15-JUN-2002) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA
On Jun 15, 2002 this sequence version replaced gi:17017552.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.zoo@hri.nih.gov
----- Project Information
Center project name: crq
Center clone name: 504H16

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BASE COUNT 53762 a 38501 c 36655 g 49899 t 400 others
ORIGIN

Query Match 2.9%; Score 41; DB 2; Length 179217;
Best Local Similarity 50.2%; Pred. No. 2.5;
Matches 101; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

QY 843 TACCTGTCTGCTTCTGCTGTACACAGGACTACGAGCTACCGCTGCGGAGGAGC 902
```

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

## ----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 177994 bases at least Q40  
Consensus quality: 178622 bases at least Q30  
Consensus quality: 178774 bases at least Q20  
Insert size: 136000; agarose-fp  
Insert size: 178817; sum-of-contigs  
Quality coverage: 12.94x in Q20 bases; agarose-fp  
Quality coverage: 9.85x in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 5 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.

\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

\* 1 32315 32414: gap of unknown length  
\* 32315 32414: contig of 32314 bp in length  
\* 66137 66136: contig of 33722 bp in length  
\* 66137 66236: gap of unknown length  
\* 66237 80633: contig of 14397 bp in length  
\* 80634 92703: gap of unknown length  
\* 80734 92700: contig of 11967 bp in length  
\* 92701 92800: gap of unknown length  
\* 92801 179217: contig of 86417 bp in length.

FEATURES  
Source

1..179217  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone="RP42-504H16"  
/clone\_lib="RP42"

## misc\_feature

1..32314  
/note="assembly\_fragment  
clone\_end:SP6  
vector\_side:left"

## misc\_feature

32415..66136  
/note="assembly\_fragment"

## misc\_feature

66237..80633  
/note="assembly\_fragment"

## misc\_feature

80734..92700  
/note="assembly\_fragment"

## misc\_feature

92801..179217  
/note="assembly\_fragment  
clone\_end:T7  
vector\_side:right"

## misc\_feature

164910..179217  
/note="clone overlaps with GenBank Accession Number  
AC13228 clone RP42-67B23 (center project name crl)"

BASE COUNT  
ORIGIN

Query Match 2.9%; Score 41; DB 2; Length 179217;  
Best Local Similarity 50.2%; Pred. No. 2.5;  
Matches 101; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

Db 106670 TACTAGCACCTTTCCGAGGTACAAACACCGTCTCTTGCTGCTGCTGCTGCTGC 106729  
Qy 903 TGCTACCCTGGCTAAGCAGTGTAAATCGCTTGCTCAGCAACCGCTATCGCTTCGG 962  
Db 106730 TGCTAAGTCGCTTCAAGTCGTCGCCAGCTGTGTGACATCCATAGAAGCAGCCGCCAGG 106789  
Qy 963 AGCTACCACTAGCTCATCTCGACACCGAGTGTCTGAACATGTGCTGAACCTTCTACTT 1022  
Db 106790 CTCCCCATCCCTGGGATTCTTCAGCAAGAACACTGGAGTGGTTGCCATTTCTCTCTC 106849  
Qy 1023 CGACGGAACAACACTTCCAGGC 1043  
Db 106850 CAATGCATGAAGCTGAAAAGC 106870

```

RESULT 11
AF374552      591 bp      mRNA      linear      PLN 27-JUL-2001
LOCUS        Thalassiosira weissflogii isolate CCMP 1336 clone 13 sexually
DEFINITION   induced protein SIG 1 mRNA, partial cds.
ACCESSION    AF374552
VERSION      AF374552.1 GI:14586994
KEYWORDS     .
ORGANISM     Thalassiosira weissflogii.
              Thalassiosira weissflogii
              Eukaryota; stramenopiles; Bacillariophyta; Coscinodiscophyceae;
              Thalassiosirophyceidae; Thalassiosirales; Thalassiosiraceae;
              Thalassiosira.
REFERENCE    1 (bases 1 to 591)
AUTHORS      Amrúst,E.V. and Galindo,H.M.
TITLE        Rapid evolution of a sexual reproduction gene in centric diatoms of
              the genus Thalassiosira
JOURNAL      Appl. Environ. Microbiol. 67 (8), 3501-3513 (2001)
MEDLINE      21365169
PUBMED       11472926
REFERENCE    2 (bases 1 to 591)
AUTHORS      Amrúst,E.V. and Galindo,H.M.
TITLE        Direct Submission
SUBMITTED    (27-APR-2001) Oceanography, University of Washington, Box
              357940, Seattle, WA 98195-7940, USA
FEATURES     Location/Qualifiers
              source          1..591
                        /organism="Thalassiosira weissflogii"
                        /isolate="CCMP 1336"
                        /db_xref="taxon:67004"
                        /clone="13"
                        /country="USA: Long Island, NY"
                        <1..>591
                        /codon_start=2
                        /product="sexually induced protein SIG 1"
                        /protein_id="AAK70420.1"
                        /db_xref="GI:14586995"
                        /translation="SHDMQCIDRNFGDGLSLRTCPGRAHVDTPRGDLDSLITGDH
NDILLAGTLPIYGITDGFFPLMSGTAGTVIANAHYMFCNSKNGLCDRLNGLCLELPG
YDGAACQRACSPCKXANTSPGSGERSNTNFVFNANSFAHRAEVQRDOCSCHG
TCMTIEQLAFLDHGNTYDLWDKDTMGCKDPGYTGP"
BASE COUNT   152 a 146 c 144 g 149 t
ORIGIN
Query Match      2.8%; Score 39.4; DB 8; Length 591;
Best Local Similarity 54.5%; Pred: No. 4.8;
Matches          79; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

Qy 278 TGAAGTGTCTCGTGGAAACGGTATATCGCTGGAGAGCTACCGACTACGCTGCTATCATCA 337
Db 198 TGTCCGGTACTGCTGGAACGGTATATCGCTAACACGGCTCATGACTACATGGAATGTCCA 257
Qy 338 CCGAGTGTGTGAAGTGTGCGATCAACTTCTACAACGAGAACCTCTTAACCTTCAACGCTG 397
Db 258 ATAAGGATTATCGCATCGCATCAATGGCGTTTGGCAATGTCTTCTGCTATGACGGTG 317
Qy 308 GAGCTTCTAGCTGTACCGCTTGCC 422

```

```

Db      318  CAGTTGCCAGCGTGCATCATGTCC 342
          |||||      |||  |  |||||
RESULT 12
AF374492      643 bp      DNA      linear      PLN 27-JUL-2001
LOCUS      Thalassiosira weissflogii isolate CCMP 1336 clone 9 sexually
DEFINITION      induced protein SIG 1 gene, partial cds.
ACCESSION      AF374492
VERSION      AF374492.1 GI:14336347
KEYWORDS
SOURCE      Thalassiosira weissflogii.
ORGANISM      Thalassiosira weissflogii
Eukaryota; stramenopiles; Bacillariophyta; Coscinodiscophyceae;
Thalassiosirophyceae; Thalassiosirales; Thalassiosiraceae;
Thalassiosira.
REFERENCE      1 (bases 1 to 643)
AUTHORS      Armbrust,E.V. and Galindo,H.M.
TITLE      Rapid evolution of a sexual reproduction gene in centric diatoms of
the genus Thalassiosira
APPL. Environ. Microbiol. 67 (8), 3501-3513 (2001)
JOURNAL      21365169
MEDLINE      11472926
PUBMED
REFERENCE      2 (bases 1 to 643)
AUTHORS      Armbrust,E.V. and Galindo,H.G.
TITLE      Direct Submission
JOURNAL      Submitted (27-APR-2001) Oceanography, University of Washington, Box
357940, Seattle, WA 98195-7940, USA
FEATURES
location/Qualifiers
1..643
/organism="Thalassiosira weissflogii"
/isolate="CCMP 1336"
/db_xref="taxon:67004"
/clone="ng"
/country="USA; Long Island, NY"
join(<1..56,141..>643)
/product="sexually induced protein SIG 1"
join(<1..56,141..>643)
/codon_start=2
/product="sexually induced protein SIG 1"
/protein_id="AAK61134.1"
/db_xref="GI:14336348"
/translation="SHDMCTCDRNFQAGDCSLRTPFGKAHVDTPRGDLASLTIGDH
NDIIAGPTLYPIGITEGIVPLMSDTAGTAVIANTADHYMECSNKGCLDRINGLCRLPG
YDGAACRSCFPSCSKANYSTPSSGQGERSTNFKVFNANSAFHGRAADVQDQCSCRG
TCMTIEQLAFLDHGNTYDLWPKDVTM"
BASE COUNT      166 a 162 c 147 g 168 t
ORIGIN
Query Match      2.7%; Score 38.6; DB 8; Length 643;
Best Local Similarity 48.0%; Pred. No. 8.2; Indels 0; Gaps 0;
Matches 110; Conservative
Qy 194 CCTGTACCCCTTGTCTTCAGAAAGACGCTCGAGCTCAGCCTAACCCCTCCTGTACCG 253
Db 198 CCCTTAATATGGCGACCACACACATATTTAGCTGTGCGACTCTTATCCATAG 257
Qy 254 CTAACCTGGTGACCCAGCTGTAAAGTGTCTCTGTTGAACCGGTATCGTGGAGGAG 313
Db 258 GTATCCGGAAGGACTTCTCTTTGATTCGGATACTGCTGGAACCGTATCGCTTAACACG 317
Qy 314 CTACCGACTAGCTGCTATCATCACCAGGTGTGAACTGTGCGCATCACTTCTACAACG 373
Db 318 CTCATGACTACATGAATGTTCCTAATAGGATATATCCGATCGCATCAATGCGCTTTCG 377
Qy 374 AGAAGCGTCCCTAACTTCAACCGCTGGAGCTTCTACCTGTACCGCTTGTC 422
Db 378 AATGTCCTCCTGGCTATGACGCGTGCAGCTTCCAGCGTGCATCATGTGTC 426
RESULT 13
HSCOLP.DGFG/C

```

LOCUS	HSCOLPDGF	16814 bp	DNA	linear	PRI 12-MAR-1997
DEFINITION	H. sapiens DNA sequence of COL1A1 gene fused with intron 1 of PDGFB gene.				
ACCESSION	X98705				
VERSION	X98705.1				
KEYWORDS	COL1A1 gene; collagen alpha 1 type I; PDGFB gene; platelet-derived growth factor beta; translocation breakpoint.				
SOURCE	Homo sapiens.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	Simon, M.P., Pedoutour, F., Sirvent, N., Grosgeorge, J., Minoletti, F., Colindre, J.M., Terrier-Lacombe, M.J., Mandahl, N., Craver, R.D., Blin, N., Sozzi, G., Turc-Carel, C., O'Brien, K.P., Kedra, D., Franss, I., Guilbaud, C., and Dumanski, J.P.				
TITLE	Deregulation of the platelet-derived growth factor B-chain gene via fusion with collagen gene COL1A1 in dermatofibrosarcoma protuberans and giant-cell fibroblastoma				
JOURNAL	Nat. Genet. 15 (1), 95-98 (1997)				
MEDLINE	97141927				
PUBMED	8988177				
REFERENCE	2 (bases 1 to 16814)				
AUTHORS	O'Brien, K.P.				
TITLE	Direct Submission				
JOURNAL	Submitted (21-JUN-1996) K. P. O'Brien, Dept. of Molecular Medicine, Clinical Genetics Unit, Karolinska Hospital, Building L-6, S-171 79 Stockholm, SWEDEN				
COMMENT	Related sequences: J03559, M10627, X07884, M20789 and K01228.				
FEATURES	Location/Qualifiers				
source	1..16814				
	/organism="Homo sapiens"				
	/isolate="cloned in cosmid pWE15 from tumor DNA from dermatofibrosarcoma protuberans (tumor TNM-1241-90)"				
	/db_xref="taxon:9606"				
	/clone_lib="cosmid D4"				
	/notes="Chimaeric cosmid clone spanning a translocation break between COL1A1 gene fused with intron 1 of PDGFB gene"				
source	1..15822				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/chromosome="17"				
	/map="q21.3-q22"				
source	15831..16814				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/chromosome="22"				
	/map="q13"				
TATA_signal	2456..2460				
gene	2484..15561				
mRNA	Join(2484..2705,4168..4362,4501..4535,4638..4673,4764..4865,5588..5659,5887..5931,6090..6143,6306..6359,6858..6911,7028..7081,7421..7474,7563..7607,7724..7777,7892..7936,8115..8168,8426..8524,8612..8656,8760..8858,8990..9043,9261..9368,9464..9517,9643..9741,9907..9960,10049..10147,11045..11098,11242..11295,11399..11452,11564..11617,12068..12112,12206..12304,12602..12709,13168..13275,13496..13549,13712..13765,13983..14090,14179..14232,14359..14412,14553..14714,14816..14923,15081..15188,15296..15349,15454..>15561)				
	/gene="COL1A1"				
	/product="collagen type I alpha 1"				
exon	2484..2705				
	/gene="COL1A1"				
	/number=1				
CDS	Join(2603..2705,4168..4362,4501..4535,4638..4673,4764..4865,5588..5659,5887..5931,6090..6143,6306..6359,6858..6911,7028..7081,7421..7474,7563..7607,7724..7777,7892..7936,8115..8168,8426..8524,8612..8656,8760..8858,8990..9043,9261..9368,9464..9517,9643..9741,9907..9960,10049..10147,11045..11098,11242..11295,11399..11452,11564..11617,12068..12112,12206..12304,12602..12709,13168..13275,13496..13549,13712..13765,13983..14090,14179..14232,14359..14412,14553..14714,14816..14923,15081..15188,15296..15349,15454..>15561)				
	/gene="COL1A1"				
	/product="collagen type I alpha 1"				
exon	2484..2705				
	/gene="COL1A1"				
	/number=1				
CDS	Join(2603..2705,4168..4362,4501..4535,4638..4673,4764..4865,5588..5659,5887..5931,6090..6143,6306..6359,6858..6911,7028..7081,7421..7474,7563..7607,7724..7777,7892..7936,8115..8168,8426..8524,8612..8656,8760..8858,8990..9043,9261..9368,9464..9517,9643..9741,9907..9960,10049..10147,11045..11098,11242..11295,11399..11452,11564..11617,12068..12112,12206..12304,12602..12709,13168..13275,13496..13549,13712..13765,13983..14090,14179..14232,14359..14412,14553..14714,14816..14923,15081..15188,15296..15349,15454..>15561)				
	/gene="COL1A1"				
	/product="collagen type I alpha 1"				
exon	2484..2705				
	/gene="COL1A1"				
	/number=1				
CDS	Join(2603..2705,4168..4362,4501..4535,4638..4673,4764..4865,5588..5659,5887..5931,6090..6143,6306..6359,6858..6911,7028..7081,7421..7474,7563..7607,7724..7777,7892..7936,8115..8168,8426..8524,8612..8656,8760..8858,8990..9043,9261..9368,9464..9517,9643..9741,9907..9960,10049..10147,11045..11098,11242..11295,11399..11452,11564..11617,12068..12112,12206..12304,12602..12709,13168..13275,13496..13549,13712..13765,13983..14090,14179..14232,14359..14412,14553..14714,14816..14923,15081..15188,15296..15349,15454..>15561)				
	/gene="COL1A1"				
	/product="collagen type I alpha 1"				
exon	2484..2705				
	/gene="COL1A1"				
	/number=1				
CDS	Join(2603..2705,4168..4362,4501..4535,4638..4673,4764..4865,5588..5659,5887..5931,6090..6143,6306..6359,6858..6911,7028..7081,7421..7474,7563..7607,7724..7777,7892..7936,8115..8168,8426..8524,8612..8656,8760..8858,8990..9043,9261..9368,9464..9517,9643..9741,9907..9960,10049..10147,11045..11098,11242..11295,11399..11452,11564..11617,12068..12112,12206..12304,12602..12709,13168..13275,13496..13549,13712..13765,13983..14090,14179..14232,14359..14412,14553..14714,14816..14923,15081..15188,15296..15349,15454..>15561)				
	/gene="COL1A1"				
	/product="collagen type I alpha 1"				
exon	2484..2705				
	/gene="COL1A1"				
	/number=1				
CDS	Join(2603..2705,4168..4362,4501..4535,4638..4673,4764..4865,5588..5659,5887..5931,6090..6143,6306..6359,6858..6911,7028..7081,7421..7474,7563..7607,7724..7777,7892..7936,8115..8168,8426..8524,8612..8656,8760..8858,8990..9043,9261..9368,9464..9517,9643..9741,9907..9960,10049..10147,11045..11098,11242..11295,11399..11452,11564..11617,12068..12112,12206..12304,12602..12709,13168..13275,13496..13549,13712..13765,13983..14090,14179..14232,14359..14412,14553..14714,14816..14923,15081..15188,15296..15349,15454..>15561)				
	/gene="COL1A1"				
	/product="collagen type I alpha 1"				
exon	2484..2705				
	/gene="COL1A1"				
	/number=1				
CDS	Join(2603..2705,4168..4362,4501..4535,4638..4673,4764..4865,5588..5659,5887..5931,6090..6143,6306..6359,6858..6911,7028..7081,7421..7474,7563..7607,7724..7777,7892..7936,8115..8168,8426..8524,8612..8656,8760..8858,8990..9043,9261..9368,9464..9517,9643..9741,9907..9960,10049..10147,11045..11098,11242..11295,11399..11452,11564..11617,12068..12112,12206..12304,12602..12709,13168..13275,13496..13549,13712..13765,13983..14090,14179..14232,14359..14412,14553..14714,14816..14923,15081..15188,15296..15349,15454..>15561)				
	/gene="COL1A1"				
	/product="collagen type I alpha 1"				
exon	2484..2705				
	/gene="COL1A1"				
	/number=1				
CDS	Join(2603..2705,4168..4362,4501..4535,4638..4673,4764..4865,5588..5659,5887..5931,6090..6143,6306..6359,6858..6911,7028..7081,7421..7474,7563..7607,7724..7777,7892..7936,8115..8168,8426..8524,8612..8656,8760..8858,8990..9043,9261..9368,9464..9517,9643..9741,9907..9960,10049..10147,11045..11098,11242..11295,11399..11452,11564..11617,12068..12112,12206..12304,12602..12709,13168..13275,13496..13549,13712..13765,13983..14090,14179..14232,14359..14412,14553..14714,14816..14923,15081..15188,15296..15349,15454..>15561)				
	/gene="COL1A1"				
	/product="collagen type I alpha 1"				
exon	2484..2705				
	/gene="COL1A1"				
	/number=1				
CDS	Join(2603..2705,4168..4362,4501..4535,4638..4673,4764..4865,5588..5659,5887..5931,6090..6143,6306..6359,6858..6911,7028..7081,7421..7474,7563..7607,7724..7777,7892..7936,8115..8168,8426..8524,8612..8656,8760..8858,8990..9043,9261..9368,9464..9517,9643..9741,9907..9960,10049..10147,11045..11098,11242..11295,11399..11452,11564..11617,12068..12112,12206..12304,12602..12709,13168..13275,13496..13549,13712..13765,13983..14090,14179..14232,14359..14412,14553..14714,14816..14923,15081..15188,15296..15349,15454..>15561)				
	/gene="COL1A1"				
	/product="collagen type I alpha 1"				
exon	2484..2705				
	/gene="COL1A1"				
	/number=1				
CDS	Join(2603..2705,4168..4362,4501..4535,4638..4673,4764..4865,5588..5659,5887..5931,6090..6143,6306..6359,6858..6911,7028..7081,7421..7474,7563..7607,7724..7777,7892..7936,8115..8168,8426..8524,8612..8656,8760..8858,8990..9043,9261..9368,9464..9517,9643..9741,9907..9960,10049..10147,11045..11098,11242..11295,11399..11452,11564..11617,12068..12112,12206..12304,12602..12709,13168..13275,13496..13549,13712..13765,13983..14090,14179..14232,14359..14412,14553..14714,14816..14923,15081..15188,15296..15349,15454..>15561)				
	/gene="COL1A1"				
	/product="collagen type I alpha 1"				
exon	2484..2705				
	/gene="COL1A1"				
	/number=1				
CDS	Join(2603..2705,4168..4362,4501..4535,4638..4673,4764..4865,5588..5659,5887..5931,6090..6143,6306..6359,6858..6911,7028..7081,7421..7474,7563..7607,7724..7777,7892..7936,8115..8168,8426..8524,8612..8656,8760..8858,8990..9043,9261..9368,9464..9517,9643..9741,9907..9960,10049..10147,11045..11098,11242..11295,11399..11452,11564..11617,12068..12112,12206..12304,12602..12709,13168..13275,13496..13549,13712..13765,13983..14090,14179..14232,14359..14412,14553..14714,14816..14923,15081..15188,15296..15349,15454..>15561)				
	/gene="COL1A1"				
	/product="collagen type I alpha 1"				
exon	2484..2705				
	/gene="COL1A1"				
	/number=1				

6306..6359  
/gene="COL1A1"  
/number=9  
6360..6857  
/gene="COL1A1"  
/number=9  
6858..6911  
/gene="COL1A1"  
/number=10  
6912..7027  
/gene="COL1A1"  
/number=10  
7028..7081  
/gene="COL1A1"  
/number=11  
7082..7420  
/gene="COL1A1"  
/number=11  
7421..7474  
/gene="COL1A1"  
/number=12  
7475..7562  
/gene="COL1A1"  
/number=12  
7563..7607  
/gene="COL1A1"  
/number=13  
7608..7723  
/gene="COL1A1"  
/number=13  
7724..7777  
/gene="COL1A1"  
/number=14  
7778..7891  
/gene="COL1A1"  
/number=14  
7892..7936  
/gene="COL1A1"  
/number=15  
7937..8114  
/gene="COL1A1"  
/number=15  
8115..8168  
/gene="COL1A1"  
/number=16  
8169..8425  
/gene="COL1A1"  
/number=16  
8426..8524  
/gene="COL1A1"  
/number=17  
8525..8611  
/gene="COL1A1"  
/number=17

Query Match 2.7%; Score 38.6; DB 9; Length 16814;  
Best Local Similarity 46.7%; Pred. No. 11;  
Matches 122; Conservative 0; Mismatches 139; Indels 0; Gaps 0;  
QY 440 GAGCTCTGACCGCTGGAACGCTGTACCATCTGGCTCAGTGTAACTGGCTGTGCTTA 499  
DB 8568 GAGGTTAGAAAGTGGCAAGGAGACACTGAGTCGGGGACACTTACAGAGCCGCGG 8509  
QY 500 CCGGAACCGCTCTGGACGAGGAGTGACACCGACTAGTCGCTCTTTTCAACGAGTGTG 559  
DB 8508 CCAGAGGCGCAGGGGCGGCGGCTCACCACGACACCTTACAGAGCCGCGGCGG 8449  
QY 560 TGAAGTGTCCGCTGACCTTCTACTACACGGAACACCGGAAACACCCCTTCAACCTG 619  
DB 8448 GGCCCTTGGGACACGAGCTTCAACCTGGAATCAGAGAAAGGACATATCAGAGCCGCTG 8389  
QY 620 GAAAGTCTAGTGTACCCCTTCTGCTATCAAGCCTGTAACTGGCTCAGGCTACCC 679

Db 8388 GGAAACCCAACTTGCCTCTCGGATGCGAGGAGGAGGAGGAGGAGGAGGAGCA 8329  
QY 680 TGGGAAACGAGCTACCATCA 700  
DB 8328 ATGATCAGACTCTAAGATCA 8308  
RESULT 14  
AP003600/c 341950 bp DNA linear BCT 28-NOV-2001  
LOCUS Nostoc sp. PCC 7120 plasmid pCC7120alpha DNA, complete genome,  
DEFINITION section 1/2.  
ACCESSION AP003600 BA000020  
VERSION AP003600.1 GI:17135539  
KEYWORDS  
SOURCE Nostoc sp. PCC 7120 plasmid:pCC7120alpha DNA.  
ORGANISM Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
REFERENCE  
AUTHORS Kaneko,T., Nakamura,Y., Wolk,C.P., Kuritz,T., Sasamoto,S.,  
Watanabe,A., Iriiguchi,M., Ishikawa,A., Kawashima,K., Kimura,T.,  
Kishida,Y., Kohara,M., Matsumoto,M., Matsuno,A., Muraki,A.,  
Nakazaki,N., Shimo,S., Sugimoto,M., Takazawa,M., Yamada,M.,  
Yasuda,M. and Tabata,S.  
TITLE Complete genomic sequence of the filamentous nitrogen-fixing  
cyanobacterium Anabaena sp. strain PCC 7120  
JOURNAL DNA Res. 8 (5), 205-213 (2001)  
MEDLINE 21595285  
REFERENCE 2 (bases 1 to 341950)  
AUTHORS Kaneko,T.  
TITLE Direct Submission  
JOURNAL Submitted (02-MAY-2001) Takakazu Kaneko, Kazusa DNA Research  
Institute, The First Laboratory for Plant Gene Research; Yana  
1532-3, Kisarazu, Chiba 292-0812, Japan  
(E-mail:kaneko@kazusa.or.jp,  
URL:http://www.kazusa.or.jp/cyanobase//,  
Tel:81-438-52-3935(ex.2338), Fax:81-438-52-3934)  
FEATURES  
source  
1. .341950  
/organism="Nostoc sp. PCC 7120"  
/db\_xref="taxon:103690"  
/plasmid="pCC7120alpha"  
/note="synonym:Anabaena sp. PCC7120"  
complement(165..548)  
/gene="all7002"  
complement(165..548)  
/note="orf7002"  
/codon\_start=1  
/transl\_table=11  
/product="transposase"  
/protein\_id="BAB78086.1"  
/db\_xref="GI:17135540"  
/translation="MWRKSYPTDLTDMWEILSPILPPAKEGSHPTTDMREICNAIY  
YHLKTCQWNMLPGDFPSPSYVYTRKWKQKGVWEKLNHTLRGQVRAKLCKSTQPSA  
IAADSQSVKTAERKGMCTALMELKR"  
1123..1719  
/gene="alr7003"  
1123..1719  
/gene="alr7003"  
/note="orf\_ID:alr7003"  
/codon\_start=1  
/transl\_table=11  
/product="transposase"  
/protein\_id="BAB78087.1"  
/db\_xref="GI:17135541"  
/translation="MEQVNTETKVHLEVTANNLSIAEANKSPVWPRFPONTSLK  
RKVQSAQAQARRDIIHARVWSLRISGLSVQAIQSLGVAKNTVNYLRSTFTERRER  
SDKGLSLNPNYHDYLLSRWNSGNHSTOKLFEELRAVGYTGSYATVSRFTYRLKTPGF  
PPAKGRKNASLRVSSSSHRPLTPSRVTALVLRREPLI"  
complement(1734..2477)  
/gene="all7004"  
complement(1734..2477)



